

Sampling and analysis of iron, steel and other ferrous metals

Part 5. Guidelines on statistical procedures

ICS 77.080

Confirmed November 2008

NO COPYING WITHOUT BSI PERMISSION EXCEPT AS PERMITTED BY COPYRIGHT LAW

BSI
Documents



Document Sharing Hub

Committees responsible for this British Standard

The preparation of this British Standard was entrusted to Technical Committee ISE/18, Sampling and analysis of iron and steel, upon which the following bodies were represented:

- BCIRA
- British Steel Industry
- Department of Trade and Industry (Laboratory of the Government Chemist)
- Ministry of Defence
- The Welding Institute

This British Standard, having been prepared under the direction of the Engineering Sector Board, was published under the authority of the Standards Board and comes into effect on 15 July 1997

© BSI 1997

The following BSI references relate to the work on this standard:
Committee reference ISE/18
Draft for comment 94/304216 DC

ISBN 0 580 26455 6

Amendments issued since publication

Amd. No.	Date	Text affected

Contents

	Page
Committees responsible	Inside front cover
Foreword	v
Section 1. General	
1.1 Scope	1
1.2 References	1
1.3 Definitions	1
Section 2. Theoretical considerations	
2.1 Accuracy, trueness and precision	2
2.2 Distribution of errors	3
Section 3. Principles of statistical evaluation	
3.1 Degrees of freedom	11
3.2 The <i>t</i> -test	11
3.3 Repeatability and reproducibility limits <i>r</i> and <i>R</i>	11
3.4 Within-laboratory reproducibility limit <i>R_w</i>	12
3.5 Standardization testing and precision testing	12
3.6 Regression analysis	12
Section 4. Experimental design	
4.1 Planning	13
4.2 Standardization tests	13
4.3 Precision tests	16
4.4 Reporting of results	16
4.5 Data truncation	16
4.6 Rounding of results	17
4.7 Choice of test samples	17
Section 5. Analysis of variance: method 1	
5.1 Effect of replication	18
5.2 Nomenclature	18
5.3 Analysis of variance applied to an example	19
5.4 Calculation of variance and standard deviation from accepted results	22
5.5 Tolerance	22
Section 6. Analysis of variance: method 2	
6.1 Direct mathematical method	23
6.2 Summary of results from a test programme	23
6.3 Calculation of sums of squares: alternative procedure	23
6.4 Degrees of freedom	24
6.5 Table of variances	24
6.6 Calculation of repeatability and reproducibility	25
6.7 Statement of precision	25
Section 7. Analysis of variance: method 3	
7.1 Simplified method using data transformation	26
7.2 Example using transformed data	26
7.3 Recovery of statistical results from transformed data	26

	Page
Section 8. Intermediate precision	
8.1 Introduction	28
8.2 Intermediate precision conditions	28
8.3 Estimation of time-different intermediate precision	28
8.4 Formation of the data table	29
8.5 Calculation of sums and squares	30
8.6 Separating the sums of squares	30
8.7 Preparation of analysis of variance table, ANOVA	31
8.8 Summary of analysis of variance	32
Section 9. <i>F</i>-variance ratio test	
9.1 Theoretical considerations	33
9.2 Applications for the <i>F</i> -statistic	33
9.3 Examples of the use of the <i>F</i> -statistic	34
Section 10. The <i>t</i>-test	
10.1 General introduction	37
10.2 Comparison of two means	37
10.3 Confidence limits, confidence interval	38
10.4 Comparison of two methods	40
10.5 Method of checking the correct application of a procedure	42
10.6 Application to regression analysis	42
Section 11. Tests for outliers and stragglers	
11.1 Theoretical considerations	43
11.2 Cochran's test for excessive variance	43
11.3 Grubbs' test for outlying single results or laboratory means	44
Section 12. Regression and correlation	
12.1 Regression analysis	46
12.2 Examples of regression calculations	46
12.3 Correlation coefficient	47
12.4 Confidence limits of regression lines	48
12.5 Probability limits for a regression line	50
Section 13. Graphical presentation	
13.1 95 % confidence limits of regression lines	51
13.2 New statistical tests	52
13.3 Youden diagrams	54
Section 14. Analytical limits	
14.1 Uncertainty factors for estimated standard deviations	57
14.2 Practical measurements of the <i>r/R</i> ratio	58
14.3 Limit of detection	58
14.4 Choice of factor <i>k</i> for limit of detection	58
14.5 Limit of determination	60
14.6 Precision of determination	61
Section 15. Tests for a normal distribution	
15.1 Introduction	62
15.2 Non-normal distributions	62
15.3 Use of normal probability graphs	62
15.4 Example of a normal probability plot	64

	Page
Annexes	
A (normative) Summary of statistical terms and symbols	66
B (informative) Statistical tables	71
Tables	
1 Nickel content determinations	3
2 Theoretical probabilities	6
3 Convergence of standard deviation	8
4 Distribution of 1000 normal standard deviates	10
5 Relationship between degrees of freedom and t	11
6 Example of rounding rules	17
7 Results for variance analysis	18
8 Laboratory averages, standard deviations and totals	20
9 Collated results for 11 laboratories ($p = 11$)	21
10 Evaluations from collated results	21
11 Statistical summary	22
12 Manganese content of five alloys	23
13 Sample C manganese	23
14 ANOVA analysis of variance	24
15 Precision data for five manganese alloys	25
16 The use of data transformations	27
17 Comparison of transformed data with original unmodified results	27
18 Data and statistical totals	29
19 ANOVA table (offset vanadium content)	31
20 F -ratios	34
21 Mean, repeatability and reproducibility from two standard methods	35
22 Variances and numbers of degrees of freedom	36
23 Ratios for F -test	36
24 Manganese content of steel bars	37
25 Offset equivalents for manganese content results	37
26 Determination of vanadium content by flame atomic absorption spectrometry	39
27 Significance of differences in pairs of means	40
28 Standard deviations and squares for Cochran's and Grubbs' tests	43
29 Cobalt content in steel in ascending order	44
30 Repeatability limits for vanadium content	46
31 Statistical parameters for regression analysis	47
32 Repeatability limits, predicted from regression equations	47
33 Minimum correlation against probability	48
34 Repeatability limits at 95 % probability based on logarithmic regression	50
35 Values of the uncertainty factors for estimates of repeatability and reproducibility standard deviations (and of the corresponding repeatability and reproducibility limits)	57
36 Probability statistics for transformed results	64
B.1 Critical values for the t -test (two-tailed)	71
B.2 Critical values for the F -variance ratio test (one-tailed) at the 0.05 significance level	72

	Page
B.3 Critical values for the F -variance ratio test (one-tailed) at the 0.01 significance level	73
B.4 Critical values for the F -variance ratio test (two-tailed) at the 0.05 significance level	74
B.5 Critical values for the F -variance ratio test (two-tailed) at the 0.01 significance level	75
B.6 Critical values for Cochran's test	76
B.7 Critical values for Grubbs' test	77
B.8 Indicators for Mandel's $\underline{h}$ and $\underline{k}$ statistics at the 1 % significance level	78
B.9 Indicators for Mandel's $\underline{h}$ and $\underline{k}$ statistics at the 5 % significance level	79
Figures	
1 Results of 57 determinations of 3.19 % nickel	4
2 Illustration of tendency of histogram to approach normal curve as intervals decrease and replicates increase	5
3 Ratio of area between limits y_1 and y_2 to total area	6
4 Distribution of 1000 random normal deviates	9
5 Uncertainty of precision measure s_r at 0.95 probability	14
6 Uncertainty of precision measure s_R at 0.95 probability	15
7 Linear and logarithmic regression lines; confidence limits	49
8 Logarithmic relationship between vanadium content (% (m/m)) and repeatability, r , or reproducibility, R and R_w	51
9 Mandel's $\underline{h}$ and $\underline{k}$ statistics for precision data for vanadium	53
10 Vanadium analysis data	55
11 Youden diagram, two-sample plot	56
12 Frequency distribution for means two standard deviations apart	59
13 Frequency distribution for means 3.29 standard deviations apart	60
14 The calibration curve, decision limit, detection limit and capability of detection	61
15 Density functions and normal probability graphs of cumulative distributions	63
16 Vanadium analyses plotted on normal probability graph paper	65
List of references	Inside back cover

Foreword

This Part of BS 6200 has been prepared by Technical Committee ISE/18.

BS 6200 is a multipart British Standard covering all aspects of the sampling and analysis of iron, steel and other ferrous metals. A general introduction is given in BS 6200 : Part 1 : 1991, which lists the contents of all the existing and projected Parts, Sections and Subsections.

This Part of BS 6200 supersedes BS 4237 : 1967, which is withdrawn. BS 4237 was based upon a report prepared on behalf of the British Iron and Steel Research Association Methods of Analysis Committee by an ad hoc panel which has been set up to examine the difficulties experienced in obtaining and expressing adequately the reproducibility of procedures used in ferrous metallurgical analysis. The report related particularly to the methods of analysis originally published as Parts of BS 1121 and later incorporated in BSI Handbook No. 19 : 1970, which has now been replaced by BS 6200.

This Part of BS 6200 revises and considerably expands the contents of BS 4237, taking into account the latest developments in the application of statistical procedures to methods of analysis. It relates in particular to the provisions of BS ISO 5725. BS ISO 5725-1, -2, -3, -4 and -6 have been published; BS ISO 5725-5 is in preparation.

This Part of BS 6200 is not intended to replace the general standards on statistics, but to complement them by explaining the theory and illustrating with practical worked examples the use of statistical tests in relation to the methods of analysis given in BS 6200 : Part 3 (and associated BS ENs) for the benefit of those without expert statistical knowledge.

Compliance with a British Standard does not of itself confer immunity from legal obligations.

Summary of pages

This document comprises a front cover, an inside front cover, pages i to vi, pages 1 to 80, an inside back cover and a back cover.

Section 1. General

1.1 Scope

This Part of BS 6200 gives guidance on the statistical procedures used during the development of methods for the analysis of iron and steel, in particular those methods described in BS 6200 : Part 3 and associated BS ENs. It explains the significance of the statistical data given in these standards. This Part of BS 6200 also gives guidance on the planning of standardization and precision tests and covers some aspects of the preparation and use of certified reference materials.

1.2 References

1.2.1 Normative references

This Part of BS 6200 incorporates, by dated or undated reference, provisions from other publications. These normative references are made at the appropriate places in the text and the cited publications are listed on the inside back cover. For dated references, only the edition cited applies; any subsequent amendments to or revisions of the cited publication apply to this Part of BS 6200 only when incorporated in the reference by amendment or revision. For undated references, the latest edition of the cited publication applies, together with any amendments.

1.2.2 Informative references

This Part of BS 6200 refers to other publications that provide information or guidance. Editions of these publications current at the time of issue of the standard are listed on the inside back cover, but reference should be made to the latest editions.

1.3 Definitions

For the purposes of this Part of BS 6200, the definitions given in BS ISO 5725 and in BS ISO 3534-1 : 1993 apply.

NOTE. For ease of reference, a summary of the statistical terms and symbols used in this Part of BS 6200 is given in annex A.

Section 2. Theoretical considerations

2.1 Accuracy, trueness and precision

2.1.1 Trueness and precision

The terms trueness and precision are used to describe the accuracy of a measurement method (see A.38 and A.23).

Trueness refers to the closeness of agreement of the average of a large number of test results to the true result, though in practice the absolute result is seldom known, and the accepted reference value may be substituted (see A.1).

Precision refers to the closeness of agreement between test results.

2.1.2 Accuracy

The general term accuracy is now used to refer to trueness and precision combined. It is a measure of the displacement of a test result from the reference value due to a combination of random and systematic errors (see A.2).

NOTE. The term bias (see A.6) has been in use for statistical matters for a very long time and is still used. Because it causes certain philosophical objections among members of some professions (such as medical and legal practitioners), the positive aspect is now emphasized by the use of the term trueness.

2.1.3 Trueness assessment

There is no mathematical means of assessing whether values obtained are true. This has to be done by analysing samples or synthetic solutions made from elements, or compounds of known formulae, of high purity and in which the amount of the element to be determined is accurately known. Interfering elements should be absent from all reagents.

One of the following alternative methods may be selected.

- a) A standardized method may be agreed to yield a true value.
- b) A reference material may be given an accepted value by consensus.

2.1.4 Reference material

A reference material is defined as a material or substance one or more of whose property values are sufficiently homogeneous and well established to be used for:

- a) the calibration of an apparatus;
- b) the assessment of a measurement method;
- c) for assigning values to materials.

NOTE. A reference material can be in the form of a pure or mixed gas, liquid or solid. Examples are pure elements or compounds as solids or in solutions of accurately known concentrations used for calibration in chemical analysis.

2.1.5 Certified reference material

A certified reference material, which is accompanied by a certificate, has one or more properties certified by a procedure which establishes its traceability to an accurate realization of the unit in which the property values are expressed, and for which each certified value is accompanied by an uncertainty at a stated level of confidence.

Testing for stoichiometry should always be part of the development of a new method where the primary aim is accuracy.

2.1.6 Repeatability limit

The term repeatability limit r is used to denote the statistical result derived from interlaboratory precision experiments using the same method. It is often called repeatability and was formerly known as repeatability value.

In the normal and correct use of the method, for one experienced operator using the same apparatus, reagents etc. and repeating the method within a short period of time, the absolute difference between two results on identical material should not exceed the repeatability limit r in more than one case in twenty, on average.

It is a measure of the consistency of immediate replicate determinations made by following the standardized procedure (see A.28).

2.1.7 Reproducibility limit

The term reproducibility limit R is used to denote the absolute difference between single results obtained by operators in different laboratories. It is often called reproducibility and was formerly known as reproducibility value.

In the normal and correct operation of the same procedure using identical material, the reproducibility limit should not be exceeded by more than one case in twenty, on average. It is a measure of the probable maximum difference of single results between laboratories using the standardized procedure (see A.32).

2.1.8 Effects of trueness and precision

It is stressed that repeatability and reproducibility limits alone carry no information on the accuracy of values obtained. Thus a method of analysis can lack trueness (and show bias towards either high or low results), and still be repeatable, have good precision, but be inaccurate. Alternatively, the presence of an interfering element in a particular sample can result in a departure from trueness by giving precise but inaccurate (biased) results.

In worst cases, lack of precision can prevent the recognition of lack of trueness. On the other hand the results of a method can show wide scatter, and therefore poor precision, but the mean of a very large number of replicates can be close to the true result.

In practice, accuracy, repeatability and reproducibility should be taken into account, and ideally both trueness and precision are sought. Once it has been established that there is no detectable departure from trueness, then better precision (as measured by repeatability and reproducibility limits) is accompanied by a narrowing of the range which is expected to include the true value.

2.2 Distribution of errors

2.2.1 Range, mean, median and mode

2.2.1.1 General theory

This brief account gives the statistical theory underlying the consideration of possible ways of defining repeatability and reproducibility limits.

Analysts know that when replicate determinations are made on a sample, the results are frequently not identical even if the sample is homogeneous. Errors occur in, for example:

- a) weights, balances and during weighing;
- b) the preparation and dispensing of solutions;
- c) photometric measurements;
- d) wavelength settings during optical measurements.

Small errors are more frequent than large ones and can be systematic or random. The systematic errors contribute to lack of trueness and the random errors to scatter and lack of precision.

The statistical analysis usually assumes trueness is satisfactory, unless the precision experiment is designed to test it. It also assumes that errors are equally likely to be positive or negative and normally distributed in magnitude.

In general there is a tendency for some errors to neutralize the effects of others by operating in the reverse direction and it is usually found that in replicate determinations, results in which the random errors almost neutralize one another occur more frequently than those in which they are additive. It is the combination of these errors that appears in the final analytical figure.

2.2.1.2 Range

An illustration of the typical condition described in 2.2.1.1 is given in figure 1, in which the results of 57 determinations of the quantity of nickel in a nominal 3.19 % (m/m) nickel steel (see table 1) are plotted at 0.01 % intervals and the frequency of results at each level shown as ordinates.

The 57 results have a range from the lowest to the highest result of 3.10 % to 3.25 % (see A.25).

Table 1. Nickel content determinations

No.	Result % (m/m)	No.	Result % (m/m)	No.	Result % (m/m)
1	3.20	20	3.21	39	3.18
2	3.19	21	3.22	40	3.20
3	3.19	22	3.19	41	3.19
4	3.20	23	3.18	42	3.18
5	3.20	24	3.18	43	3.14
6	3.17	25	3.10	44	3.21
7	3.20	26	3.16	45	3.25
8	3.18	27	3.20	46	3.18
9	3.19	28	3.13	47	3.19
10	3.21	29	3.19	48	3.20
11	3.18	30	3.19	49	3.20
12	3.20	31	3.18	50	3.17
13	3.22	32	3.20	51	3.21
14	3.18	33	3.20	52	3.17
15	3.17	34	3.19	53	3.22
16	3.17	35	3.21	54	3.18
17	3.20	36	3.16	55	3.18
18	3.16	37	3.23	56	3.19
19	3.21	38	3.21	57	3.19

2.2.1.3 Mean and median

One way of summarizing the results of this experiment would be to quote the mean result, defined as the sum total of all the results divided by the number of results.

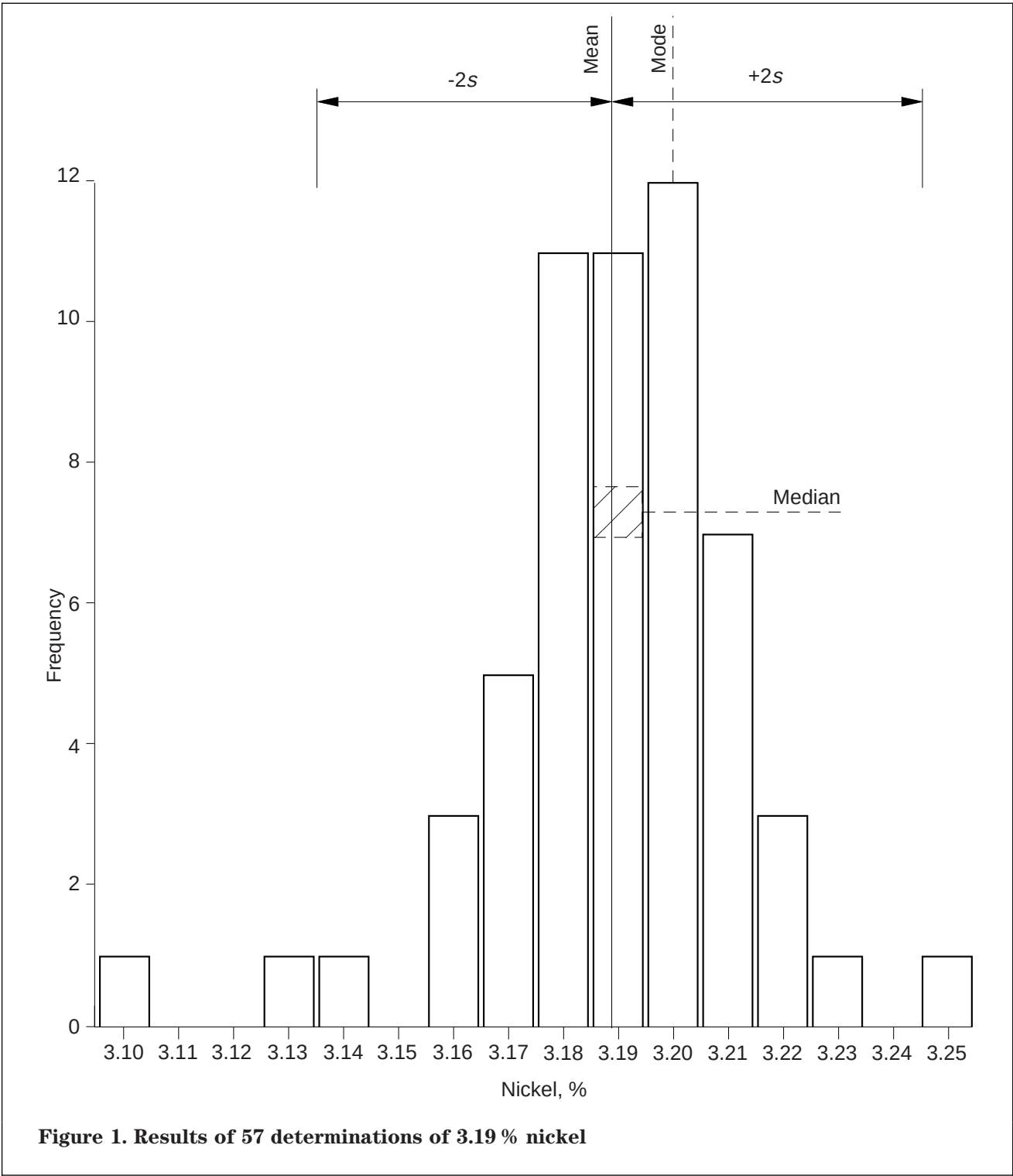
$$\begin{aligned}\text{Mean result} &= 181.78/57 \\ &= 3.189 \% (m/m) \text{ nickel}\end{aligned}$$

The mean gives no information on the scatter of results.

When there are relatively few results, one result differing widely from the remainder can cause a large difference between the mean and the value for the mean obtained in an extended trial.

In such cases the median result, which is the middle value, can give a better prediction of the average of the series which would be obtained if it were extended by further measurements. It circumvents the problem of dealing with possible outliers because it is unaffected by their actual values. This is of particular advantage when the number of results is small so that the mean would be strongly affected (see A.19).

In figure 1 the median is the twenty ninth result from either extreme, and this falls within the 3.19 % nickel column. There is no indication of spread, and, in the absence of three decimal place listings in the results reported, a more precise estimate of the median cannot be made.



2.2.1.4 Mode

The results can be characterized by their mode. The mode is the most frequent value, in this case 3.20 %, which occurs 12 times in the 57 results. There can be one or more modes in a data set (see A.20).

If the results are symmetrically distributed about the mean, then in a large population of results the values of mean, median and mode will coincide. However when analytical results are near to the lower limit of determination, negative results are removed on logical grounds and the distribution tends to have a tail towards the higher end. In such cases a single mode is numerically lowest, with the median somewhat greater, and the mean value highest. Similar suppression of anomalous values can occur when attempting determinations at concentrations approaching 100 %.

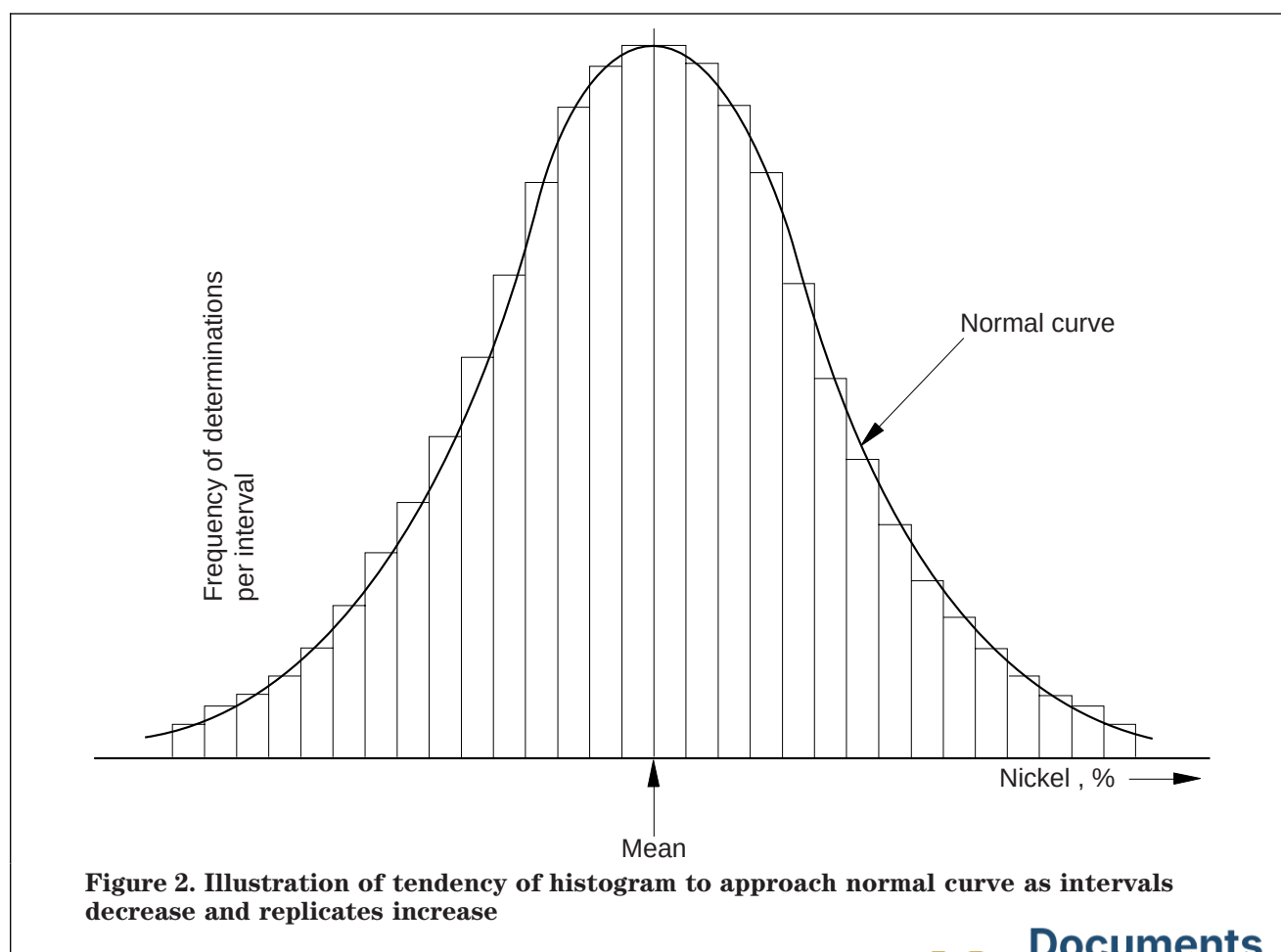
2.2.2 Normal distribution

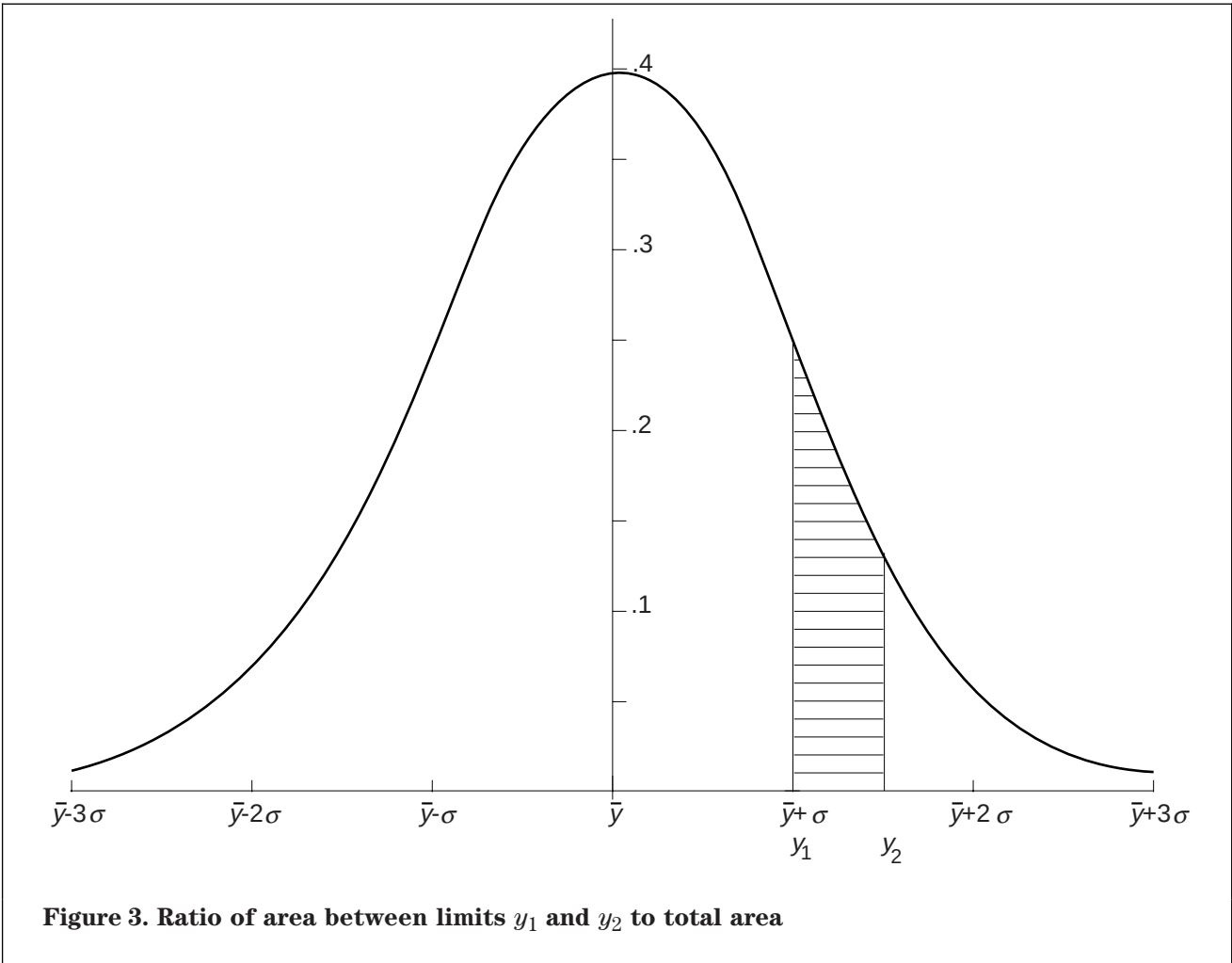
2.2.2.1 It is important to consider the change that would occur in figure 1 if the intervals were decreased in size and at the same time the number of determinations increased. Under these conditions the steps between ordinates would become smaller until ultimately an infinitesimally small interval size together with an infinity of results would yield a smooth

frequency distribution curve as shown in figure 2. This curve (known as the normal curve or Gaussian distribution) (see A.21), has several properties which are of immediate interest.

2.2.2.2 Firstly it should be noticed that the tails (extremities of the curve) are asymptotic to the x -axis. There are no discontinuities which could help in defining repeatability or reproducibility and all that can be inferred is that results become less frequent as their net positive or negative errors increase. There is a chance, albeit a slight one, that net errors of considerable magnitude will occasionally arise and therefore the range values of a set of replicate determinations will tend to increase with the number of replicates (see A.25).

2.2.2.3 A second important property of the normal curve is that the proportion of results occurring within any interval on the x -axis is equal to the area bounded by the x -axis, the interval lines and the normal curve (see figure 3) divided by the total area under the curve. The normal curve therefore permits the calculations of the probability of a single determination lying within any chosen limits.





2.2.2.4 In figure 3:

- a) y_1 lies at $\bar{y} + \sigma$;
- b) y_2 lies at $\bar{y} + 1.5 \sigma$.

The total area under the normal curve is unity. Between the axis, normal curve and y_1 or y_2 the areas are 0.3413 and 0.4332 respectively, so the area shaded is 0.0919 of the total. In a series of normally distributed results the probability of a single result falling in the shaded area is 9.19 %.

2.2.2.5 Because most analytical results with which this Part of BS 6200 is concerned have an approximately normal distribution, the fact that the tails of the normal curve are asymptotic signifies that any reproducibility limits given for an analytical method have to be based on a statement of probability. Thus as the reproducibility limits are increased the probability of falling outside them decreases but never vanishes. Some agreed probability level (usually 95 %) has therefore to be clearly stated with any quoted repeatability or reproducibility level. Since the

probability of any result falling within these limits is not 100 % an occasional analysis may be expected to lie outside them. The theoretical probabilities, expressed as percentages, of a single result from a population having a normal distribution of errors falling within

$\text{mean} \pm k\sigma$;

where

$k = 1 \text{ to } 4$

are given in table 2.

Table 2. Theoretical probabilities		
k	$\bar{y} \pm k\sigma$	Probability
1	$\bar{y} \pm \sigma$	68.26
2	$\bar{y} \pm 2\sigma$	95.46
3	$\bar{y} \pm 3\sigma$	99.74
4	$\bar{y} \pm 4\sigma$	99.994

2.2.3 Variance

2.2.3.1 The dispersion of results caused by random errors about the mean is measured by the variance of the population. This is obtained by summing the squares of the differences between each result and the overall mean, and then adding the squares. The grand total is divided by the number of results to give the variance of the population (see A.41). This is given by the equation:

$$\text{Variance} = \frac{\sum (y_i - \bar{y})^2}{n} \quad (1)$$

where

$\bar{y}$ is the mean of n values;

y_i represents an individual value;

n is the number of results in the entire population.

2.2.3.2 The procedure given in 2.2.3.1 applies only if a measurement has been made on every member of the population. As this is usually impracticable, an estimated variance is calculated by using a divisor which is one less than the number of measurements to allow for the use of only a sample (see 3.1 and A.41). This is given by the equation:

$$\text{Estimated variance} = \frac{\sum (y_i - \bar{y})^2}{(n - 1)} \quad (2)$$

where

$\bar{y}$ is the mean of n values;

y_i represents an individual value;

n is the number of results in the sample.

2.2.3.3 The mathematical treatment, as shown in equations 1 and 2, gives a greater weighting to the results further from the mean, but these are fewer than the results close to the mean. Variance therefore gives a balanced view of the set.

2.2.3.4 Each source of random error, for example those arising from a change of analyst, instrument, calibration, laboratory, etc. makes its own variance contribution to the overall scatter. The variances are additive. From the results of a carefully designed experimental program the sources of error can be made separable and quantifiable, and perhaps minimized or eliminated. Tests for normal distribution are discussed in section 15.

2.2.4 Standard deviation

2.2.4.1 The positive square root of the variance is called the standard deviation of the set of measurements. It is denoted by the symbol σ if every possible measurement has been made, and refers to the dispersion of the entire n units in the population of results. It is given by the equation:

$$\sigma = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n}} \quad (3)$$

2.2.4.2 When only a sample is measured the standard deviation is an estimate, in which case it is denoted by

s . The estimated standard deviation s of n measurements is calculated from the equation:

$$s = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n - 1}} \quad (4)$$

where

$\bar{y}$ is the mean of n values;

y_i represents each individual value in turn;

s is an estimate of the standard deviation derived from n results with $n - 1$ degrees of freedom (see 3.1 and A.35).

2.2.4.3 To obtain the difference of each value from the mean by squaring these differences and then taking their sum is laborious. It is usually more convenient to sum the numbers and their squares and use the alternative and equivalent equation 5 which is suited for calculators:

$$s = \sqrt{\frac{\sum (y_i^2) - (\sum y_i)^2/n}{n - 1}} \quad (5)$$

and derive the mean separately from

$$\bar{y} = \sum y_i / n \quad (6)$$

NOTE. For detailed instructions on the economical application of formulae using calculators or computers see section 7.

2.2.5 Relative merits of range and standard deviation

2.2.5.1 Range

Both range and standard deviation are statistical measures of the spread of replicate chemical analyses. Range is the difference between the greatest and least values observed. Its only real advantages are the ease with which it can be calculated and its obvious interpretation. It has two serious disadvantages:

- There is no indication of the limits to the values a variable can take. All that can be said is that high and low values are likely to be less frequent than the intermediate ones. Consequently range is very sensitive to the occurrence of an extreme value.
- Range is liable to fluctuate considerably from batch to batch and tends to increase with the number of measurements.

2.2.5.2 *Standard deviation*

The main advantages of standard deviation are:

- a) It is possible to estimate the proportion of values which lie within chosen limits, see 2.2.2.
- b) It is relatively stable, regardless of the number of values considered (above a certain minimum) or the incidence of extreme values. It is also possible to eliminate results which have a specified probability of being erroneous (outliers) and to draw attention to stragglers, using consistent statistical tests.
- c) It is possible to express standard deviation as a quantity made up of components due to laboratory error, operator error and random error; by the use of a good experimental plan. Conversely, since it is possible to separate these individual sources of error the major weaknesses of a measurement procedure can be readily identified.

2.2.5.3 *Comparison of range with standard deviation*

2.2.5.3.1 The comparison between range and standard deviation can be illustrated by an example. A block of 1000 values was taken from a table of 25 000 random normal deviates having a mean of 3.000 and a standard deviation of 0.100 (see reference [1]).

The first five values were taken and the range and standard deviation computed. A further five values were added and the range and standard deviation re-computed. This process was repeated until results were obtained for 5, 10, 20, 50, 100, 500 and 1000 values as shown in table 3.

Table 3. Convergence of standard deviation			
No. of values	Mean	Range	Standard deviation
<i>n</i>	$\bar{y}$	$y_{\max} - y_{\min}$	<i>s</i>
5	3.012	0.23	0.094
10	3.013	0.24	0.083
20	3.010	0.28	0.092
50	3.002	0.39	0.095
100	2.988	0.40	0.091
500	2.995	0.59	0.094
1000	2.998	0.71	0.099

Figure 4 shows the histogram of the 1000 values.

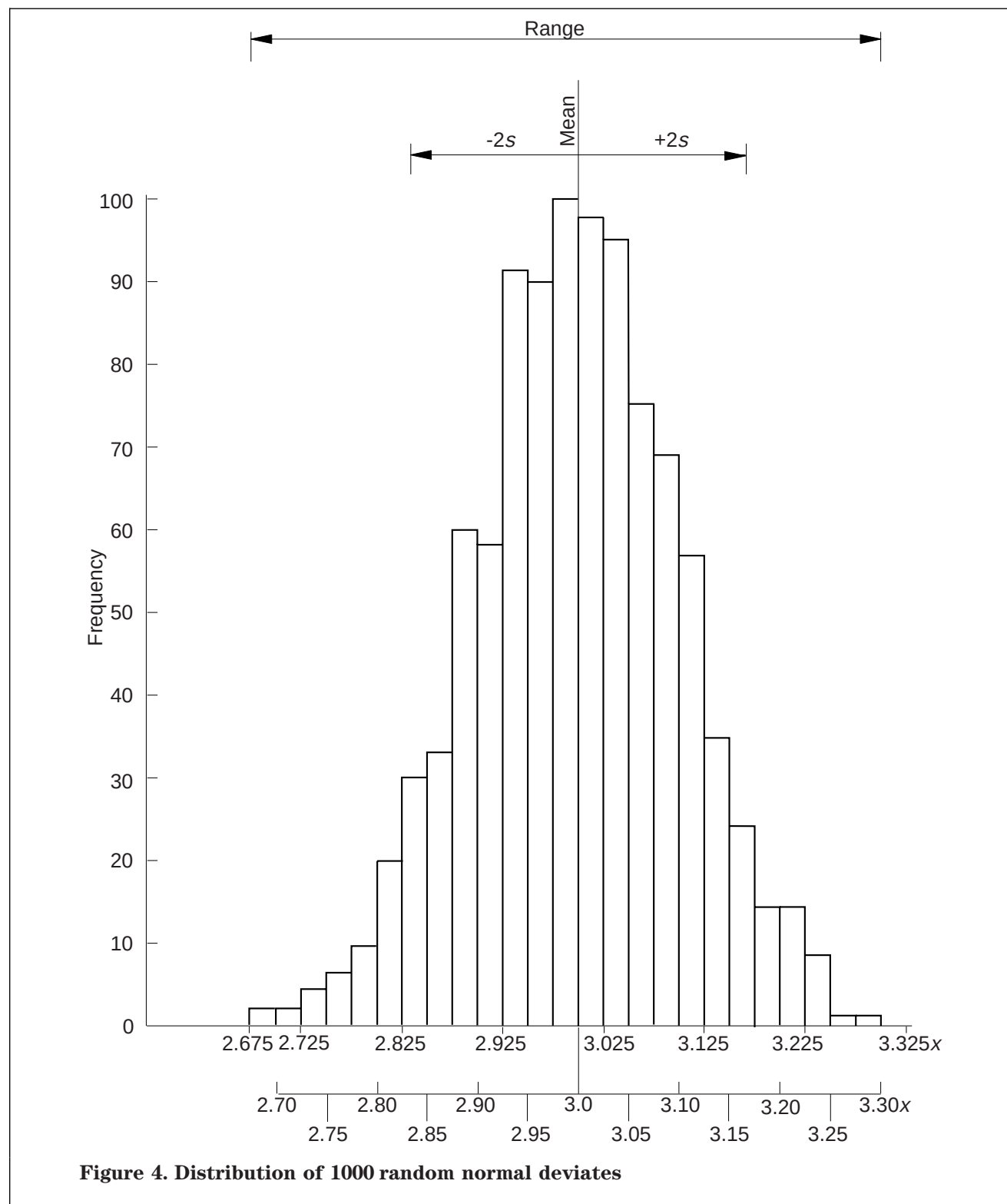
2.2.5.3.2 If a set of results is distributed normally the reproducibility of results at any chosen probability can be calculated.

For a large sample:

- a) 68.26 % of results lie within ± 1 standard deviation of the mean;
- b) 95.46 % of results lie within ± 2 standard deviations of the mean;
- c) 99.74 % of results lie within ± 3 standard deviations of the mean.

For practical purposes these probability statements are usually approximated to two thirds, once in twenty and twice per thousand respectively.

2.2.5.3.3 Table 3 shows that as more figures become available the mean and standard deviation remain fairly stable and close to their true values, whilst the range steadily increases as the probability of exceptionally large or small values increases. The inherent stability of standard deviation, in spite of extremes of values, can be seen.



2.2.5.3.4 Table 4 compares the experimental results with the theoretical distribution in terms of standard deviation.

Table 4. Distribution of 1000 normal standard deviates				
Range	Lower limit	Upper limit	Expected percentage	Observed percentage
			%	%
Mean $\pm 1s$	2.9	3.1	68.26	67.8
Mean $\pm 2s$	2.8	3.2	95.46	95.7
Mean $\pm 3s$	2.7	3.3	99.74	99.7

The range increases with the number of tests and cannot be utilized for prediction in the same manner.

This principle is illustrated using the 57 results of analysis for nickel listed in table 1 and plotted in figure 1.

These were all individual determinations made at different times. The mean value of the 57 results was 3.189 and the standard deviation 0.0239.

81 % of the results lie with $\pm 1s$ and 93 % within $\pm 2s$, and all but one result within $\pm 3s$. The discrepancies between predicted and observed distribution almost certainly are caused by excessive rounding before statistical evaluation. This is discussed in 4.5 and 4.6.

2.2.5.3.5 The normal curve is an ultimate response which is never attained in practice because the number of replicate determinations is always finite. The problem of inferring the characteristics of the normal curve, which represents an infinity of results from a finite number of replicate determinations, is always present. Means and standard deviations from experimental data are estimates of the true mean and standard deviation (i.e. of the normal curve); as the number of results increases so the estimates approach more closely to the true values. If only a few results are available the mean and standard deviation can be poor estimates of the true values derived from the results of all possible or a much larger number of measurements (see section 14).

2.2.6 Uncertainty

2.2.6.1 At present there is no agreed definition of uncertainty. Two differing approaches are being advocated. They agree that it is important to distinguish between error and uncertainty.

2.2.6.2 Error is defined as the test result minus the true or accepted value.

2.2.6.3 Uncertainty is an estimate attached to a test result which characterizes the range of values within which the true value is asserted to lie. The exact or true value is usually unknowable, and the final measured result is still only an estimate because of random errors and imperfect corrections for systematic effects.

2.2.6.4 Two measures of uncertainty have been proposed.

a) Standard uncertainty sums all the variances of random error and the square root of the sum is taken. Expanded uncertainty multiplies this result by a factor $k = 2$ for general use to cover the whole range with a probability close to unity.

NOTE. This definition is recommended for easy and standardized use in practical cases.

b) Strict uncertainty adds the variances only of those components which are truly random, and to the square root of this sum it then adds half the range of those uncertainty components which are systematic, for example method bias.

Section 3. Principles of statistical evaluation

3.1 Degrees of freedom

When making statistical calculations from a class of results, for a given data set, the grand total is fixed by the known results, so when the individual results are entered, the last result is forced without any choice.

For example, in a calculation using results from analyses in triplicate ($n = 3$), any two determinations can be entered, but the third is forced by the requirement to have the results yield the experimentally observed mean value for that set of three. The set has only two degrees of freedom (see A.14).

Similarly for results in triplicate from eleven laboratories, ($p = 11$), the total number of degrees of freedom for all 33 results is $(pn - 1) = 32$.

The 11 within-laboratory mean results use two degrees of freedom each so

$$\text{degrees of freedom} = p(n - 1) = 22$$

i.e. within-laboratory results require a total of 22.

This leaves ten degrees of freedom for between-laboratory results which is given by

$$\text{degrees of freedom} = (p - 1) = 10$$

This also gives the correct overall total.

Using the correct number of degrees of freedom is an essential step in the analysis of variance and in probability calculations.

3.2 The *t*-test

3.2.1 For a very large and normally distributed set of results, it can be asserted that individual results have a 95 % probability of falling within the range given by $(\text{mean} \pm 1.96 \text{ standard deviations})$. There is only a 5 % probability that this assumption is not true. The significance of the hypothesis is said to be 0.05. However, for a small number of results the standard deviation is an estimate of the population value. The penalty for the lack of data is an increased uncertainty in both the mean and the calculated standard deviation. Both can differ slightly from the true values for the population (see A.40).

3.2.2 To compensate for the small sample number the factor of 1.96 is increased to allow a possible wider range at a given probability. The uncertainty depends upon the confidence interval or significance required and the number of results available. The appropriate correction factor t to replace 1.96 is derived from a table of t values. If only six replicates are available, with five degrees of freedom, then to obtain an estimate for the true standard deviation from a very large population of possible results distributed normally, and with a 95 % probability of it not being exceeded, the observed standard deviation has to be multiplied by $t = 2.57$ instead of $t = 1.96$.

3.2.3 Table 5 lists a series of values of t selected from a more comprehensive table. It shows that when approximately 30 or more results are available a factor of two can be used with a 95 % probability of further results falling within the predicted range.

A fuller set of values for other probabilities and numbers of replicates is found in the table B.1. Note that the table lists degrees of freedom, which usually is one less than the number of results in a class (see 3.1). There is also an uncertainty in the calculated value for the mean itself. Confidence limits of a mean are discussed in 10.3.

Table 5. Relationship between degrees of freedom and t

Degrees of freedom	t -factor for standard deviation with a 95 % confidence interval
5	2.57
10	2.23
15	2.13
20	2.09
30	2.04
60	2.00
120	1.98
∞	1.96

3.2.4 Further applications of the t -test occur:

- in calculating confidence limits of linear regression equations (see 12.4);
- in testing the statistical equality of experimental means (see 10.2);
- in checking the correct use of an analytical procedure (see 10.5);
- in calculating confidence intervals (see 10.3).

3.3 Repeatability and reproducibility limits r and R

For the reasons discussed in 3.2, the requirement to express repeatability in a simple way can be met by quoting a multiple of the standard deviation.

The standard factor of 2.83 is chosen as a multiplier for the standard deviation in order to give 95 % confidence intervals (i.e. $\bar{y} \pm 2.83s$).

This is accepted as of more practical value than a confidence interval of 68 % calculated from the mean $\pm$ one standard deviation ($\bar{y} \pm s$).

The origin of the multiplier 2.83 is explained in 5.3.5.2. Reproducibility expressed in similar terms is covered in 5.3.5.3.

3.4 Within-laboratory reproducibility limit R_w

In some of the European and ISO standards included in BS 6200 : Part 3 an additional descriptor of precision is used. This is applied to the results from single determinations made in one laboratory by the same operator, with the same equipment but on a different day and with recalibration of volumetric solutions or instruments. It is intended as a guide to the typical stability of determinations within a single laboratory estimated from the results of the group of laboratories participating in the original precision experiment. The time interval and other variables between this pair of determinations are unspecified. Therefore the value of R_w serves only as a general guide (see A.43). Its numerical value usually falls between the extreme cases of repeatability and reproducibility, as do the possible sources of error. The separation of sources of variance is discussed more fully in sections 5, 6 and 7, and intermediate precision R_w in section 8.

3.5 Standardization testing and precision testing

3.5.1 Producing a standard procedure has two stages.

a) The first stage is the standardization phase which is intended to:

- 1) discover and remove technical difficulties;
- 2) to produce a clear unambiguous text;
- 3) result in a procedure meeting a preset target for accuracy and precision.

b) The second stage uses precision tests to provide data covering the full range of the procedure in order to demonstrate that the precision obtained by experienced operators is satisfactory.

NOTE. Accuracy is usually only rechecked in so far as true results are obtained on reference materials.

3.5.2 The standard deviations of the sources of normally distributed random error are not additive, but their squares (variances) are (see 2.2.4). Therefore experiments can be designed to separate them and evaluate the relative importance of the sources.

One of the objects of the experimental design should be to assess the several sources of variance which together cause the differences in results obtained from test to test and laboratory to laboratory.

3.5.3 Even when all the known sources of error are taken into account and the same operator in the same laboratory carries out repeat tests within a short time interval, small random errors will appear in the results, measured as part of the within-operator variance.

If tests on the same sample are carried out by different operators, in general they do not obtain identical results. The spread of results for each operator can differ from the mean of the spreads from all the operators, and the difference can be measured as the between-operators variance.

Individual laboratory means can differ from the mean of all the laboratory means, and the differences can be measured as between-laboratory variance.

Such information is of great value and it is of the utmost importance that sufficient work be done in the standardization stages of method development to ensure that reliable estimates of repeatability and reproducibility limits are available to justify commencing precision testing.

In the same way, other causes of variance can be assigned, isolated, and possibly removed after suitably designed trials at the standardization stage of method testing.

3.5.4 When the separate sources of variance have been estimated, it is possible to assess the readiness of the method for precision testing and organize the necessary test programme.

The results should allow precise statements to be made about its repeatability and reproducibility limits. It is then possible to calculate with any desired degree of confidence the number of tests required to make an accurate estimates within pre-set limits. The method is explained in 14.1.

3.6 Regression analysis

3.6.1 In cases where repeatability and reproducibility limits vary with concentration level or type of sample, it is necessary to quote different confidence limits for each level or sample.

3.6.2 The relationship between repeatability or reproducibility and concentration can also be derived from the statistical procedure of regression analysis. When a method has been tested using concentrations spread over its entire working range, regression analysis gives more reliable estimates of precision than are obtained from a single level because all the results can be used to smooth out some of the experimental uncertainties in estimates at individual levels.

3.6.3 The relationship between concentration and repeatability (or reproducibility) limits is sometimes linear, but a logarithmic function is much more commonly found. It is therefore desirable to select test samples spaced at approximately logarithmic intervals (e.g. at 1, 2, 5, 10, 20, 50, 100, etc.) to cover the desired range of concentration.

NOTE. Linear and logarithmic regression are discussed in section 12.

Section 4. Experimental design

4.1 Planning

4.1.1 Much effort is expended in the search for new and improved analytical methods, proceeding through trials to standardize all the details of the procedure to assess trueness and culminating in a final precision trial to establish its repeatability and reproducibility limits. A method which satisfies the criteria of the laboratory examinations to this stage should be known to be true or, at least, be thought more accurate on theoretical grounds than any alternative method. If premature attempts are made to measure repeatability and reproducibility in a final precision trial before the method is proved to measure unambiguously the variables it is designed to measure in the standardization trials the result is a waste of time and resources.

4.1.2 Ideally it would be informative to have results from more than one analyst per laboratory in order to be able to separate within-operator, between-operator, and between-laboratory variances. However, scarcity of analytical resources usually renders this ideal impractical and in general only one highly experienced operator from each laboratory participates.

NOTE. The within-operator and between-operator-same-laboratory variances, both of which in practice could contribute errors from a single laboratory, are inseparable when using this type of experimental plan.

4.1.3 The usual practice is for each operator to carry out at least three determinations on each sample.

The statistical analysis of the results is easier if the same number of operators:

- a) each report the same number of test results on each material;
- b) work at each concentration level.

4.1.4 International cooperative work relating to the development of methods of analysis of iron and steel currently requires:

- a) the participation of at least 15 laboratories;
- b) the use of six to ten different test materials with compositions covering the entire range of the proposed method.

Additionally, in collaborative studies there is a requirement for participating laboratories to be provided by:

- 1) a minimum of three countries (for European standards);
- 2) a minimum of five countries (for ISO standards).

4.1.5 For some of the older methods in BS 6200 : Part 3 the minimum national requirement was for ten laboratories to make analyses in triplicate.

4.1.6 The relationships between number of participating laboratories and the uncertainty of estimates of repeatability and reproducibility, standard deviation and limits are summarized in figures 5 and 6 taken from BS ISO 5725-1 : 1994.

The symbol γ represents the ratio of the reproducibility standard deviation σ_R to the repeatability standard deviation σ_r and is given by the equation:

$$\gamma = \sigma_R / \sigma_r \quad (7)$$

Practical values of γ can be in the range 1 to 4, but are commonly about 1.5 to 2 for ferrous material analyses.

Further information is given in section 14 and in more detail in BS ISO 5725-1 : 1994.

4.2 Standardization tests

4.2.1 The standardization trials are used to test the suitability of a method procedure before finalising the text. During the standardization stage the samples supplied need not necessarily be certified reference materials provided that:

- a) they are homogeneous;
- b) they have stable composition;
- c) the content of the element to be determined is known with reasonable accuracy.

4.2.2 The primary objectives are to establish that the precision attainable meets the objectives of the programme and to reveal shortcomings in the procedure, for example inadequate or erroneous instructions, interferences, non-stoichiometric reactions, etc. If it is suspected that the procedure is subject to day to day variation, single determinations on each sample should be made on separate days.

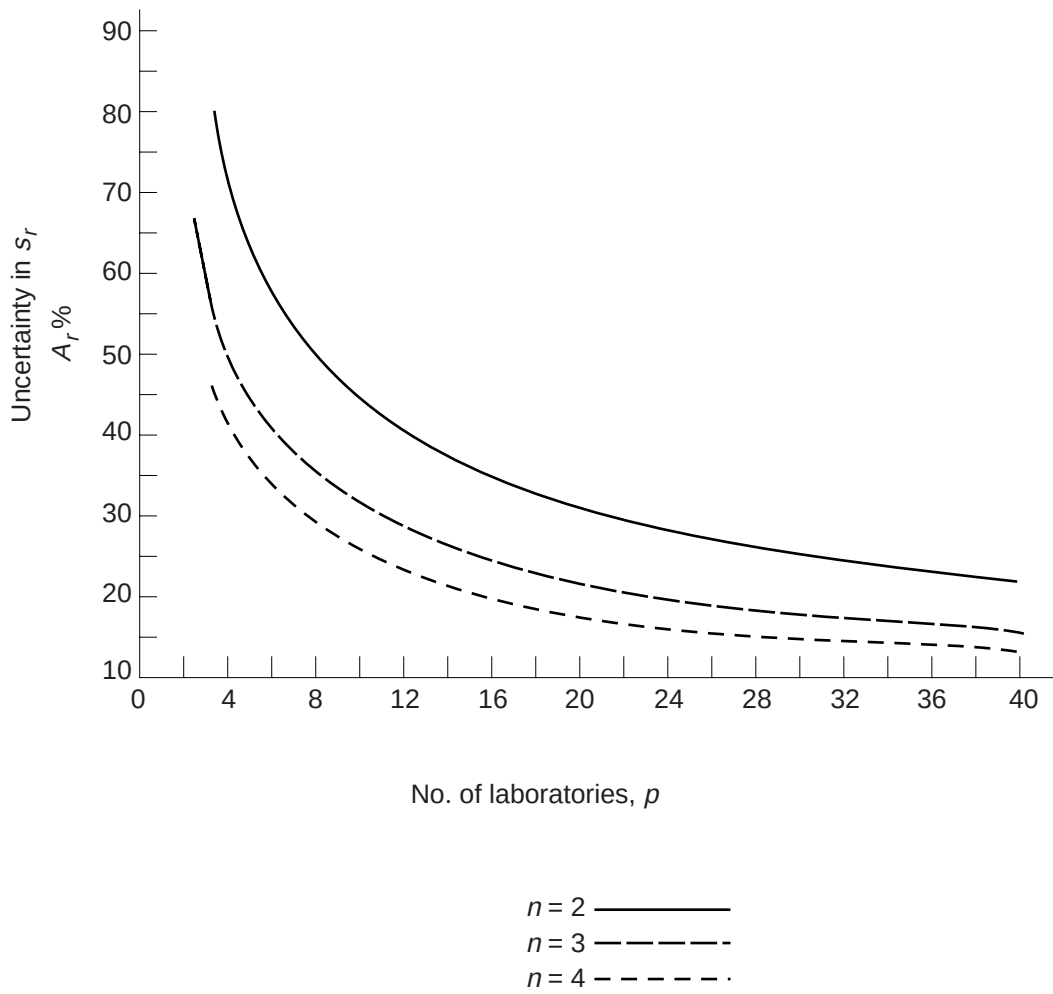


Figure 5. Uncertainty of precision measure s_r at 0.95 probability

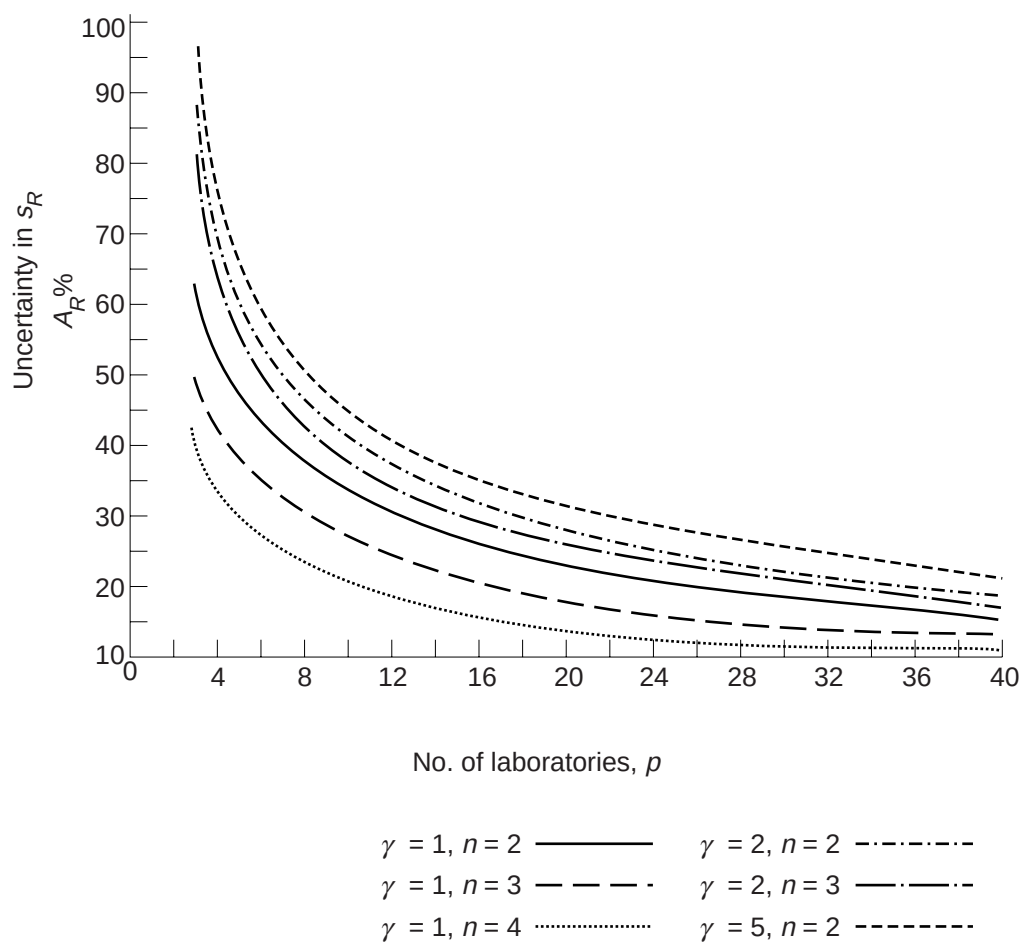


Figure 6. Uncertainty of precision measure s_R at 0.95 probability

4.3 Precision tests

4.3.1 The precision tests are applied only when a satisfactory procedure has been finalized and all the procedural details adequately tested and described.

4.3.2 Statistical analysis assumes that test results are independent. If it is considered that previous results can influence subsequent test results (and so reduce the estimate of repeatability variance) it may be necessary to provide specimens coded in such a way that the operator does not know which are supposedly identical, and to randomize the order of testing (see also **13.3**).

4.3.3 When testing for precision, the identity of the samples and their exact composition should not be given to the cooperating laboratories, and the programme should include instructions to disclose results only to the convenor or the appointed statistician.

4.3.4 The quantity of sample distributed should be sufficient only for the number of determinations requested.

4.3.5 No departures from the agreed procedure should be permitted during precision testing.

4.3.6 For planning purposes the number of laboratories to be recruited depends upon the desired probability level, repeatability and reproducibility (see section **14**).

4.3.7 Each analyst should have satisfactory experience and skills in the particular techniques to be used. However the precision experiment should not be restricted to experts in the field because that could lead to better than typical performance.

If the operator is not already familiar with the procedure from standardization trials then practice determinations should first be made on materials having compositions similar to the samples prescribed for the precision trial.

4.4 Reporting of results

4.4.1 To be able to apply the appropriate statistical procedure to analytical results it is essential that two (or more) complete and independent determinations should be made on identical material within a short time by each participating analyst, for example within one day, if the method so permits. The current pattern requires each operator to make two determinations in quick succession using the same solutions, apparatus and calibration etc., and the same operator to make a third determination on a later day again using the same apparatus, but with a new instrumental calibration and freshly standardized solutions.

4.4.2 It is not essential that the determinations at different levels of concentration should all be determined on the same occasion, but if the procedure includes a calibration, then so should each batch of determinations.

4.4.3 If day to day variations are known to be absent the complete series of tests may be carried out as a single batch provided it is practicable.

4.4.4 If the procedure requires the measurement of a blank then at least one is needed with each batch of determinations.

4.4.5 The result obtained from every test should be reported. No result should be discarded unless it is known that the procedure has not been performed satisfactorily (i.e. values will not be discarded merely because they disagree with other values). This procedure should be adhered to without exception, and the circumstances reported to the convenor or statistician.

4.5 Data truncation

4.5.1 Immediately prior to the final precision trial a decision should be reached on the number of significant figures to be retained in each observation. It is often difficult to assess the effects of rounding on intermediate observations. Therefore, a statistician should be consulted to determine the number of digits to be retained at each stage in the calculations.

4.5.2 It is the responsibility of the convenor of the working group undertaking the investigation:

- a) to ensure that all co-operators are fully aware of the dangers associated with rounding;
- b) to consult with the statistician who will analyse the data;
- c) to stipulate the number of significant figures required for each value in the final table of results. The number of digits reported should be in accordance with the remit of the experimental program, any additional digits being omitted.

4.5.3 These instructions should be followed by all the participating laboratories. The number of digits will vary according to the particular case but to obtain a reliable estimate of the standard deviation will usually require results to be reported to at least one significant figure more than is required by the procedure, and therefore more than may be customary in routine analysis.

4.5.4 It is important that the results from all operators are quoted to the same number of significant figures set out in the program or on a form supplied for reporting results. The statistical calculation depends on small but significant differences between very large numbers. Rounding can occur:

- a) of the original results by automatic instrumentation;
- b) in the calculation;
- c) by truncation in calculators or computers.

The effect of rounding can cause large errors. Therefore results should not be rounded before the end of the statistical process. (The results in table 1 were rounded, which now prevents further statistical analysis.)

4.6 Rounding of results

4.6.1 If the required precision is unobtainable, spurious zeros should not be added. If the experimental results have more significant figures than are to be reported, then any obvious roundings should be made to the required precision. In cases falling at mid-point the rounding rules given in BS 1957 : 1953 should be followed. These are summarized as follows.

- a) Apply rounding once only to any individual result. The removal of digits by a sequence of rounding steps is not permitted.
- b) Choose the rounded value (up or down) which is the nearer. If the figures to be discarded fall exactly between two prospective alternative rounded values proceed to c).
- c) When rounding values which have one more significant figure than the number specified, proceed as follows.
 - 1) Last digit is zero
Discard and take no further action.
 - 2) Last digit is 1 to 4
Discard and leave the remaining value unaltered.
 - 3) Last digit is 6 to 9
Discard and increase the penultimate digit by 1.
 - 4) Last digit is 5
If the penultimate digit is zero or even, discard the 5 and take no further action.
If the penultimate digit is odd, discard the 5 and increase the penultimate digit by 1.

4.6.2 The data in table 6 illustrate the rounding rules applied to various examples of observed values and in cases where:

- a) the values are rounded to two significant figures;
- b) the values are rounded to two decimal places.

Table 6. Example of rounding rules

Observed value	Rule applied:		Rounded to:	
	For two significant figures	For two decimal places	Two significant figures	Two decimal places
7.2547	Obvious	Obvious	7.3	7.25
7.345	Rule b	Rule c	7.3	7.34
7.555	Rule b	Rule c	7.6	7.56
7.665	Rule b	Rule c	7.7	7.66

4.7 Choice of test samples

4.7.1 General

The selection of samples for the final experimental plan should receive most careful consideration in order that the statistical analysis can isolate the various effects due to sample options.

4.7.2 Type of material

In certain cases special modifications are recommended within the method procedure to deal with the presence of possible interfering elements. Such modifications could influence the reproducibility of the results obtained. They should be examined for discrepancies before combination with the main body of results.

4.7.3 Level of content

Where possible test samples should be chosen to span the range of element levels and matrices within the scope of the method draft (see A.17). In this connection the method may stipulate definite sample masses, dissolution procedure, aliquot fractions, solution strengths, photometer cell sizes, etc., depending upon the level of content of the element to be determined. Some discretion is permitted regarding these variables in certain methods and in such cases the working group should agree on the exact conditions to be employed for each sample. These should be tabulated in the programme and with the results obtained.

Individual laboratories should not be permitted a free choice. Any alternatives within a procedure would introduce an uncontrolled variable into the trial or one of its levels.

4.7.4 Number of samples

The minimum number of samples is usually six, with ten as the maximum. Because the relationship between concentration and repeatability or reproducibility is often logarithmic it is also appropriate to choose concentrations in the test samples in a similar progression.

For example, approximately 1, 2, 5, 10, 20, 50, 100.

This ensures an even coverage of the range and improved regression analysis for a logarithmic relationship between precision and concentration.

One sample should test the lowest intended value, and another should have approximately the maximum content for the range specified in the method.

Section 5. Analysis of variance: method 1

5.1 Effect of replication

5.1.1 Sources of variance

When values obtained by more than one operator, or laboratory, are compared, additional sources of variance are introduced. Therefore a wider scatter is expected than when the same number of determinations is made by a single operator.

There can be some variation between the means obtained by different operators or laboratories. These give rise to between-laboratory or between-operator reproducibility, variance or standard deviation (see A.5).

Analysis of variance (ANOVA) is a statistical procedure which facilitates the estimation of the contributions of each of the separate sources of random error to the total variance (see A.4).

5.1.2 Factors affecting variance

5.1.2.1 Before considering the manner in which it is possible to isolate the sources of variation, two important facts have to be stated. These are discussed in 5.1.2.2 and 5.1.2.3.

5.1.2.2 It is known that the mean of several replicate determinations is more reliable than an individual determination. Also the absence of bias is assumed. The reliability of the estimate of the mean increases with the number of determinations used in its calculation.

Similarly the standard deviation of a set of means (each calculated from the same number of observations) about the grand mean is smaller than the standard deviation of the individual results.

If s is the standard deviation of the individual results then $s/\sqrt{n}$ is the standard deviation of a mean derived from n results. To halve the standard deviation of an analytical figure, the mean of four independent determinations is required.

5.1.2.3 The second important fact is that the overall variance is equal to the sum of the individual variances contributed by the sources of random experimental errors.

5.2 Nomenclature

5.2.1 It is possible to break down the variances within the data in table 7 into the following two components:

- those arising from different laboratories;
- those due to replicate determinations by each operator.

Table 7. Results for variance analysis

Operator from laboratory	Sample 1			Sample 2		
	Test 1	Test 2	Test 3	Test 1	Test 2	Test 3
1	0.162	0.161	0.153	0.060	0.065	0.065
2	0.160	0.160	0.165	0.053	0.053	0.053
3	0.155	0.153	0.170	0.057	0.057	0.065
4	0.168	0.172	0.172	0.067	0.070	0.070
5	0.165	0.165	0.167	0.056	0.057	0.057
6	0.168	0.166	0.165	0.064	0.062	0.068
7	0.162	0.168	0.162	0.056	0.059	0.059
8	0.159	0.162	0.167	0.063	0.058	0.064
9	0.163	0.162	0.168	0.064	0.065	0.059
10	0.160	0.160	0.160	0.058	0.058	0.059
11	0.167	0.158	0.162	0.067	0.060	0.062

5.2.2 In the general case it is desirable to have the same number of replicates from each laboratory at a given level, and the nomenclature identifies n replicates from each of p laboratories from each of q levels of concentration.

- Laboratories are numbered:

$$i = 1, 2, 3, \dots p$$

- Levels are numbered:

$$j = 1, 2, 3, \dots q$$

- Replicates are numbered

$$k = 1, 2, 3, \dots n_{ij}$$

For example, k_{ij} is the k th replicate from laboratory i at level j .

5.3 Analysis of variance applied to an example

5.3.1 Data

The results in table 7, from which statistical conclusions are drawn, come from analytical values from 11 laboratories making determinations in triplicate on two samples containing differing levels of the element concerned.

The discussion which follows interleaves the steps in the statistical logic with the practical calculations, and then draws the appropriate conclusion from that step.

NOTE 1. This example uses the same statistical procedure as BS ISO 5725-2 : 1994 which is suitable for all cases, including those where the tables of results are unbalanced due to missing experimental data, or the subsequent removal of outliers, leaving some cells incomplete (see sections 8 and 9).

NOTE 2. The procedure uses a series of forms, here condensed into table 8 by working to the detailed instructions in the text which follows. This procedure could readily be adapted into a spreadsheet, with considerable saving of effort if a recomputation became necessary as a consequence of the removal of outliers. It differs from the usual mathematical treatment because it operates on mean instead of on individual results.

5.3.2 Components of variance

5.3.2.1 General

In this case two sources of variance could be separable.

5.3.2.2 Variance from replicate determinations by each operator

It is assumed in deriving the variance for this source that each operator exhibits approximately the same variance within his own replicates. This variance (termed within-operator variance) is repeatability variance, denoted by s_r^2 . The assumption is investigated using Cochran's test (see 11.2).

5.3.2.3 Variance from change of operator or laboratory

It is assumed that laboratory or operator means have a normal distribution about the overall mean and have a variance denoted by s_L^2 . The means are tested for outliers from a distribution assumed to be normal using variants of Grubbs' test (see 11.3).

NOTE 1. There are no results enabling the separation of between-laboratory and between-operator variances in this example but laboratory is used for s_L^2 rather than s_b^2 to avoid confusion.

NOTE 2. These two variances should account for all the observed departures from the precise overall mean, but because the errors are random there is a small residual variance which could also be calculated by an extension of the procedure to be described. Residual variance is assumed to be non-significant in the statistical treatment described in this Part of BS 6200.

5.3.3 Calculation of required sums and squares

The steps in the calculations are explained in words. The recommended statistical symbolism is replaced, temporarily and for clarity, by arbitrary letters in the identification of the accumulated totals needed in the compilation of table 8.

The sequence for filling in the rows of columns a) to i) in the table is as follows.

- Starting from the results from sample 1 in table 7, the mean result for each laboratory is calculated and entered in table 8, on the row for that laboratory.
- The laboratory means in a) are squared and entered.
- The number of replicates for each laboratory is entered.
- The number of replicates in c) is squared and entered.
- The standard deviation for each laboratory is calculated from the equation and entered.

$$s = \sqrt{\frac{\sum (y^2) - (\sum y)^2/n}{n - 1}} \quad (8)$$

where

s = standard deviation;

n = number of replicates in laboratory;

y = each individual result.

- The standard deviations are squared and entered.
 - The laboratory means a) are multiplied by the replication c) and entered.
 - The squared laboratory means b) are multiplied by the replication c) and entered.
 - The squared standard deviations f) are multiplied by one less than the replication c) and entered.
- $$[i] = f) \times \{c) - 1\}$$

Each column is summed to complete tables 8a and 8b.

The observed variance of the individual operator means about the grand mean is composed of s_r^2 (see 5.3.2.2) and s_L^2 (see 5.3.2.3).

The variance of the operator mean is one third of that for a single result because three determinations have contributed. That is:

- $s_r^2/3$ in this case;
- s_r^2/n for n replicates per operator in the general case.

The variance due to change in laboratory is s_L^2 .

Hence the total variance observed for the laboratory mean is $(s_r^2/3 + s_L^2)$.